



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism.

Citation for published version:

Nuthmann, A & Henderson, J 2012, 'Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism.', *Visual Cognition*, vol. 20, no. (4-5), pp. 457-494.
<https://doi.org/10.1080/13506285.2012.670142>

Digital Object Identifier (DOI):

[10.1080/13506285.2012.670142](https://doi.org/10.1080/13506285.2012.670142)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Visual Cognition

Publisher Rights Statement:

© Nuthmann, A., & Henderson, J. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, 20((4-5)), 457-494.
[10.1080/13506285.2012.670142](https://doi.org/10.1080/13506285.2012.670142)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



This is the accepted manuscript version. The final publication is available as:

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, 20(4-5), 457-494. doi: 10.1080/13506285.2012.670142
<http://dx.doi.org/10.1080/13506285.2012.670142>

Using CRISP to Model Global Characteristics of Fixation Durations in
Scene Viewing and Reading with a Common Mechanism

Antje Nuthmann¹ and John M. Henderson²

¹University of Edinburgh, UK

²University of South Carolina, USA

Keywords: scene perception, reading, eye movements, fixation durations, computational modeling, saccadic inhibition

Corresponding Author:

Antje Nuthmann

University of Edinburgh

Psychology Department

7 George Square

Edinburgh, EH8 9JZ, UK

E-Mail: Antje.Nuthmann@ed.ac.uk

Phone: +44 (0)131 650 3459

Abstract

Fixation durations vary when we read text or inspect a natural scene. Past studies suggest that this variability is controlled by the visual input available within the current fixation. The present study directly compared the control of fixation durations in reading and scene viewing in a common experimental paradigm, and attempted to account for the control of these durations within a common modeling framework using the CRISP architecture (Nuthmann, Smith, Engbert, & Henderson, 2010). In the experimental paradigm, a stimulus onset delay paradigm was used. A visual mask was presented at the beginning of critical fixations, which delayed the onset of the text or scene, and the length of the delay was varied. Irrespective of task, two populations of fixation durations were observed. One population of fixations was under the direct control of the current stimulus, increasing in duration as delay increased. A second population of fixation durations was relatively constant across delay. Additional task-specific quantitative differences in the adjustment of fixation durations were found. The pattern of mixed control of fixation durations obtained for scene viewing has been previously simulated with the CRISP model of fixation durations. In the present work, the model's generality was tested by applying its architecture to the text reading data, with task-specific influences realized by different parameter settings. The results of the numerical simulations suggest that global characteristics of fixation durations in scene viewing and reading can be explained by a common mechanism.

239 words

Introduction

During the course of a complex visual-cognitive task such as reading or picture viewing, our eyes move from one location to another at an average rate of 3 to 5 times per second (for recent reviews see Henderson, 2003, 2007; Rayner, 1998, 2009a). Between these movements (*saccades*), the eyes come to rest for brief periods of time (*fixations*). The durations of individual fixations have been found to reflect ongoing perceptual and cognitive activity, providing a powerful method for investigating underlying perceptual and cognitive processes (Rayner, 1998). Reading and scene viewing both engage vision, but reading also engages language. Despite the undeniable differences between tasks, stimulus processing in both tasks is subject to the restrictions that arise from the operation of the oculomotor system.

Fixation Durations in Reading and Scene Perception

The influence of visual and cognitive factors on fixation durations in complex visual-cognitive tasks is widely acknowledged. In reading, fixation durations are influenced by perceptual variables as well as by linguistic variables (see Rayner, 1998, for a review). Specifically, local word properties such as length, frequency, and predictability exert consistent effects on fixation durations (e.g., Kliegl, Nuthmann, & Engbert, 2006). Readers look longer at long words than at short words (e.g., Just & Carpenter, 1980; Kliegl, et al., 2006; Rayner, Sereno, & Raney, 1996). Word frequency effects represent a key empirical marker for lexical processing: Low frequency words are fixated longer than high frequency words (e.g., Inhoff & Rayner, 1986; Just & Carpenter, 1980; Kliegl, et al., 2006; Rayner & Duffy, 1986). Such benchmark findings are now explicitly accounted for by computational models of eye-movement control in reading (see below).

In comparison, the relationship between fixation durations and visual-perceptual and cognitive influences in scene viewing is less well studied. There is evidence that fixation durations are sensitive to global scene processing difficulty, manipulated as global image

degradation. For example, low-pass filtered images produce longer fixation durations than either high-pass or unfiltered images (Mannan, Ruddock, & Wooding, 1995), and reducing the luminance of a scene leads to increased fixation durations (Loftus, 1985). Whether local effects equivalent to visual and lexical effects in reading can be observed for objects in scenes is currently not well explored (see Henderson, 2011, for a review). Predictability (sentence-word consistency) has immediate effects on fixation durations in reading, but predictability effects in scenes (object-scene consistency) tends to be observed only on aggregate measures of fixation time (e.g., Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009). Furthermore, Wang, Hwang, and Pomplun (2010) recently investigated the effects of object size and linguistics-based and scene-based frequency on object fixation times. Large objects were fixated longer than small objects, and small and large objects induced very different frequency and predictability effects, but most of these effects were observed for gaze duration and total viewing time only, not for first fixation durations.

Task Comparisons

In reading, the “scene” consists of well-defined and neatly ordered word objects, and most of the time the eyes simply move from left to right through a line of text. In comparison, a picture of a real-world scene is typically much more complex and the direction and size of eye movements is much less predictable than in reading (Henderson, 2003). Analyses of fixation positions have provided some evidence that the eye-movement control system directs the eyes in terms of word units in reading (McConkie, Kerr, Reddix, & Zola, 1988; Rayner, 1979) and object units during scene viewing (Nuthmann & Henderson, 2010).

How do measures of eye movements compare across reading and scene viewing? At a basic level, eye movements can be characterized by means and variations in fixation durations and saccade lengths. It is clear that global eye-movement characteristics differ between the two tasks. Fixations durations in scene perception tend to have a longer average

duration than in reading, and the range of fixation durations is greater (Henderson & Hollingworth, 1998; Rayner, 1998). The basis for this difference is not yet clear. The distance the eyes move in scene perception is typically larger than that in reading (Henderson & Hollingworth, 1998; Rayner, 1998). Furthermore, studies measuring the visual or perceptual span in reading (McConkie & Rayner, 1975) and scene viewing (Saida & Ikeda, 1979) and/or real-world search (Parkhurst, Culurciello, & Niebur, 2000) suggest that information is taken in from a wider area in scenes than in reading. It could be that fixation durations in scene viewing are longer because more information is being taken in. Alternatively, longer fixation durations in scene viewing could be due to the need for more complex decisions about where to move next.

Another way to look at similarities and differences across viewing tasks is to study eye-movement measures at the level of the individual. If a given viewer's fixation durations and saccade amplitudes are long in one task, are they long in other tasks as well? Although an individual's fixation durations and saccade amplitudes tend to correlate across non-reading tasks (Andrews & Coppola, 1999; Castelhana & Henderson, 2008; Rayner, Li, Williams, Cave, & Well, 2007), eye-movement parameters in reading do not tend to correlate with scene viewing (Andrews & Coppola, 1999; Rayner, et al., 2007).

In sum, previous research has established that (1) means and distributions of fixation durations differ across reading and scene viewing, and (2) subjects with long/short fixations in scene viewing will not necessarily make long/short fixations in the reading task as well. These findings could be taken to suggest that eye-movement control fundamentally differs between scene viewing and reading. On the other hand, there are reasons to suppose that eye movements in the two tasks draw on the same underlying control processes. First, the neural circuitry for controlling eye movements is the same. Second, in both scene viewing and reading, eye movements are partly a consequence of visual acuity limitations. We move our

eyes to place the fovea on that part of the text or picture we want to see clearly. Therefore, the function of the movements is the same. Third, for both tasks, new information about the text or scene is brought into the information processing system during the time that the eyes are in fixation, and control of timing of the fixations may therefore be similar. Fourth, the fine details of how the cognitive system interacts with the oculomotor system are likely to differ as a function of task (Rayner, 2009a), but this interaction is subject to the restrictions that arise from the operation of the oculomotor system. Most notably, it takes a significant amount of time to program an eye movement, and this places constraints on oculomotor control. Based on these considerations, it has been suggested that eye-movement control models developed in one domain must have the potential for generalization to others (Engbert, Nuthmann, Richter, & Kliegl, 2005; Nuthmann & Engbert, 2009), and that modeling approaches should aim at a constructive convergence across scene viewing and reading (Henderson & Smith, 2009; Nuthmann, et al., 2010).

The Stimulus Onset Delay Paradigm

The *stimulus onset delay paradigm* provides a method for investigating whether and how fixation durations are controlled ‘online’ by the current visual input. It is particularly suitable for comparing tasks because it can be applied to visual stimuli of all kinds, and it selectively manipulates global stimulus processing difficulty.

Three decades ago, the *text onset delay* (TOD) paradigm was used to investigate the control of fixation durations in reading (Morrison, 1984; Rayner & Pollatsek, 1981). In these experiments, a visual mask was presented at the end of each saccade, which delayed the onset of the text, and the length of the delay was varied (between 25 and 300 ms). If fixation durations are directly controlled by the current visual text input, one would expect programming of the saccade to be delayed until text on which to base programming became visible (Rayner & Pollatsek, 1981). When the text onset delay lasted less than 150 ms,

fixation duration increased proportionally with the delay duration, indicating that fixation duration is under direct control. In the longer delay conditions, there was an additional population of fixations that ended while the mask was still present, suggesting that their preparation had already started on the previous fixation.

The stimulus onset delay paradigm has also been used to investigate the control of fixation durations in scene viewing (Henderson & Pierce, 2008; Henderson & Smith, 2009; Shioiri, 1993; van Diepen, Wampers, & d'Ydewalle, 1998). In the *scene onset delay* paradigm, participants examine photographs of real-world scenes while engaged in a viewing task (e.g., scene memorization, visual search). During the saccade prior to a pre-specified critical fixation, the scene is replaced with a mask, which delays the onset of the scene. The scene reappears after the manipulated delay period. Again, if fixation durations are directly controlled by the current visual input, programming of the next eye movement should be delayed until there is scene information present on which to base that programming. Two fixation populations were identified. One population increased with scene onset delay, whereas the second population remained relatively constant across delays (Henderson & Pierce, 2008; Henderson & Smith, 2009). The two populations of fixation durations were separated by a gap which is likely due to saccadic inhibition (Reingold & Stampe, 2002, 2004).

The qualitative pattern of results from scene viewing appears to be consistent with the results from reading. However, the details of the onset delay experiments for reading and scene viewing have differed, and a direct comparison has not yet been conducted. In the present study, we compared the two tasks directly to study qualitative and quantitative similarities and differences. In a within-subject design, participants were tested on both text reading and scene viewing, and the implementation of stimulus onset delays was identical in both tasks.

Models of Eye-Movement Control in Reading

Over the past three decades, research on the nature of eye-movement control in reading has not only generated a large body of empirical findings but has also led to the development of complex and sophisticated theoretical approaches and computational models (see Rayner, 2009b; Reichle, Rayner, & Pollatsek, 2003, for reviews). The following discussion will focus on the two most advanced models in the field: the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, et al., 2003; Reichle, Rayner, & Pollatsek, 2012 this issue; Reichle, Warren, & McConnell, 2009) and the SWIFT model (Engbert, Longtin, & Kliegl, 2002; SWIFT 2: Engbert, et al., 2005; SWIFT 3: Schad & Engbert, 2012 this issue). In particular, we will briefly discuss their implementations with regard to the control of fixation durations, and their potential for generalizability beyond reading.

At the level of saccade programming, both models share the notion that saccades are programmed in two stages. Evidence from basic oculomotor research suggests that there is an initial, labile stage that is subject to cancellation, followed by a non-labile stage in which the saccade program can no longer be cancelled (Becker & Jürgens, 1979). At the level of saccade timing, however, the two models differ quite substantially with regard to the mechanisms that control fixation durations. What triggers a new saccade program is conceptualized to be lexical processing (E-Z Reader) or a random timer (SWIFT).

According to the E-Z Reader model, eye movements are under the direct and immediate control of lexical properties of a given word (Reichle, et al., 1998). As a result, their deployment honors the serial order in which words occur (see Reichle, Liveredge, Pollatsek, & Rayner, 2009, for discussion). At the end of an early stage of a word identification process (L_1), a saccade is programmed to the next word. In sum, saccade programs are triggered by lexical processing, and the model thus implements a strong form of

direct control of fixation durations.

In comparison, the SWIFT model implements a weaker coupling between saccade timing and processing of the currently fixated word (Engbert, et al., 2002; Engbert, et al., 2005). In this model, saccade programs are not initiated by the completion of a cognitive process, but by an autonomous timer. However, lexical processing difficulty (of the currently fixated word) modulates fixation durations by inhibiting the timer so that it delays the initiation of the next saccade program. As a result, fixation durations will be lengthened, allowing additional time for lexical processing. The inhibition acts with a delay, so that any difficulty that comes with the processing of word n results not only in longer fixation on word n (an immediacy effect) but also in longer fixation on word $n+1$ (a spillover effect) (Engbert, et al., 2005). In short, difficulty in lexical processing produces delays in the initiation of the saccade program terminating the fixation.

The present study investigated the mechanisms that control when the eyes move. With regard to these “when” decisions in reading and scene viewing, the stimulus onset paradigm provides a test bed for models advocating a strong (E-Z Reader) or less strong (SWIFT) coupling between saccade timing and stimulus processing, an issue we will discuss in depth in the General Discussion.

Models of Attentional Selection in Natural Scenes

Most current computational models of scene viewing and visual search incorporate the concept of a bottom-up saliency map (in differing implementations) with or without top-down control (e.g., Itti & Koch, 2000; Navalpakkam & Itti, 2005; Parkhurst, Law, & Niebur, 2002; Torralba, Oliva, Castelano, & Henderson, 2006; Zelinsky, 2008; see Tatler, Hayhoe, Land, & Ballard, 2011, for a review). These models seek to predict fixation locations (where), but they typically ignore fixation durations (when). It is probably fair to say that these models describe spatial selection and attention shifts rather than gaze control. A notable

exception is the Target Acquisition Model (TAM) by Zelinsky (2008, 2012 this issue) which was designed to predict scan paths in visual search. The model attempts to detect the target in the scene and generates eye movements to bring a simulated fovea closer to what is currently considered to be the most likely target candidate. The model's behavior is driven by a target map representing the visual similarity between the target and the search scene. Immediately following each eye movement, the search scene is transformed to reflect the human visual system's retinal acuity limitations. An eye movement is made once a critical threshold is reached, and time is reflected in terms of a sequencing of eye movements. However, TAM does not incorporate an explicit saccade-programming module, and it currently does not explain the durations of individual fixations.

The CRISP Model

While there exist a number of advanced computational models explaining fixation durations (when?) and fixation positions (where?) in reading (for an overview of several of these models, see the 2006 special issue of *Cognitive Systems Research*), our own CRISP model is currently the only theoretical approach and computational model that was specifically developed to account for variations in fixation durations during scene viewing (Nuthmann, et al., 2010).

CRISP is a timer (C)ontrolled (R)andom-walk with (I)nhibition for (S)accade (P)lanning model. The model architecture can be summarized with three main principles. First, a random walk process generates inter-saccadic intervals and thus variations in fixation durations. Second, moment-to-moment difficulties in visual and cognitive processing can immediately inhibit (i.e., delay) saccade initiation, leading to longer fixation durations. Third, saccade programming comprises two stages: an initial, labile stage that is subject to cancellation, and a later, non-labile stage (Becker & Jürgens, 1979; Reichle, et al., 1998). In

the following section, these three modeling principles are discussed in a bit more detail.

Additional information can be found in Nuthmann et al. (2010).

 Insert Figure 1 about here

In CRISP, the saccade timer is implemented as a random walk process (cf., Ratcliff & Rouder, 1998; Reddi, Asrress, & Carpenter, 2003; Roitman & Shadlen, 2002). The random walk timing signal accumulates toward a threshold. When the threshold is reached, a new saccade program is initiated (Figure 1, Figure 2). The main parameter is the transition rate for the random walk (i.e., the elementary steps toward threshold), which determines how fast the process of saccade timing operates. The transition rate r_1 is defined as

$$r_1 = \frac{N}{t_{sac}}, \quad (1)$$

where N is the number of states the process can adopt, and t_{sac} is the mean duration of the timing signal. Specifically, the random walk is implemented as a discrete-state continuous-time Markov process with exponentially distributed waiting times between elementary transitions (for further details see Nuthmann, et al., 2010).

In the CRISP model, processing difficulty can inhibit (i.e., delay) saccade timing and programming, resulting in longer fixation durations. This can happen in two ways: (1) current processing demands modulate the random walk's transition rate, and (2) processing difficulties can lead to saccade cancellation (cf., Vergilino-Perez, Collins, & Doré-Mazars, 2004; Yang, 2009; Yang & McConkie, 2001) (inhibitory elements are marked in red in Figure 1). It is worth highlighting that conceptualizing the timer as a random walk process allows for continuous crosstalk between visual-cognitive processing and saccade timing: The random walk creates a trajectory approaching threshold over time which can be modulated at any point by visual-cognitive events.

Two simulations with the baseline model parameters from Nuthmann et al. (2010), summarized in Table 1, illustrate this principle. To mimic low processing load (“easy”), the random walk transition rate was arbitrarily increased ($r_2 = r_1 / 0.80$). Figure 2 shows one example of the timing signal (Figure 2a) and the fixation duration distribution obtained in the condition mimicking low processing load (Figure 2b); the distribution is shifted towards shorter fixation durations. In case of high processing load (“difficult”), the rate was decreased by the same factor ($r_2 = r_1 * 0.80$). Due to greater processing load, the completion threshold is reached later in time (Figure 2c as compared to Figure 2a). Increased processing demands slow down the timer. This delays the initiation of the next saccade program, and eventually leads to longer fixation durations. Accordingly, we observe a rightward-shift of the obtained fixation duration distribution towards longer fixation durations (Figure 2d).¹

 Insert Figure 2 about here

The model assumptions about saccade programming are consistent with current evidence concerning basic oculomotor control: Saccade programming is completed in different stages and saccade programs can partly overlap in time (Becker & Jürgens, 1979). Specifically, saccades are programmed in two stages: an early, labile stage that can be canceled by the initiation of subsequent saccadic programs, followed by a non-labile stage that is not subject to cancellation. The durations of these two programming stages are sampled from gamma distributions with means τ_{lab} and τ_{nlab} (Table 1). After termination of a non-labile saccade program, a saccade is executed with mean duration τ_{ex} (Table 1). The CRISP model combines autonomous timing and temporally overlapping saccade programming, a feature it shares with the SWIFT model of eye-movement control (Engbert, et al., 2002; SWIFT 2: Engbert, et al., 2005; Nuthmann & Engbert, 2009; SWIFT 3: Schad &

Engbert, 2012 this issue). As a consequence, saccade programs can be initiated before the processing of information from the current fixation. Instances where the corresponding saccade program was started before the onset of that particular fixation will lead to relatively short fixation durations, as was demonstrated in a simulation study (Nuthmann, et al., 2010). These simulations also demonstrated that the implemented saccade cancellation mechanism prolongs fixation durations.

Insert Table 1 about here

The current instantiation of CRISP models the control of fixation durations without taking fixation positions into account (Nuthmann, et al., 2010), a reasonable first step given the relative independence of “when” and “where” decisions in eye-movement control (see Findlay & Walker, 1999). In its current implementation the model does not perform an analysis of scene content. Although this is a limitation, the model provides a general computational framework for exploring the extent to which fixation durations are under perceptual and cognitive control during scene viewing. The long-term goal of the research program is to add a “where” module to the model.

Our current modeling efforts are guided by the principle of model generalizability (e.g., Pitt, Myung, & Zhang, 2002). We propose that the duration of fixations is similarly controlled across viewing tasks. Therefore, it is important to determine whether the model can capture global characteristics of fixation durations in other tasks like reading.

The Present Study

The aim of the present study was to explore the following working hypothesis: The timing and programming of saccades, and thus the control of fixation durations, is not fundamentally different in scene viewing and reading. Global characteristics of fixation

durations in both tasks can be explained by common model architecture. To be clear, the emphasis is on global eye-movement characteristics, ignoring the task-specific local effects on fixation durations that undeniably exist. To compare the control of fixation durations in reading and scene viewing, we used the stimulus onset delay paradigm to selectively manipulate global stimulus processing difficulty by delaying text or scene presentation during critical fixations (e.g., Henderson & Pierce, 2008; Rayner & Pollatsek, 1981). Based on results from previous studies applying the stimulus onset delay paradigm to texts (Morrison, 1984; Rayner & Pollatsek, 1981) and scenes (Henderson & Pierce, 2008; Henderson & Smith, 2009), we expect to see two populations of fixation durations, one that is directly controlled by the current visual stimulus, and a second one that is not. This qualitative signature should be observed in both text reading and scene viewing. In addition to these qualitative similarities we expect task-specific quantitative differences in the adjustment of fixation durations. The pattern of mixed control of fixation durations is modeled with the CRISP model of fixation durations (Nuthmann, et al., 2010), with task-specific influences realized by different parameter settings. If eye-movement control operates similarly in reading and scene viewing, we should be able to capture the main characteristics of both with the same basic modeling architecture.

Methods

Participants and Apparatus

Twelve participants (4 males; mean age = 21.5 yrs) took part in the experiment. All had normal or corrected-to-normal vision. Stimuli were presented on a 21-inch CRT monitor with a refresh rate of 140 Hz at a viewing distance of 90 cm. A chin rest was used to keep participants' head position stable. During stimulus presentation, participants' eye movements were recorded using an SR Research EyeLink 1000 eye tracker. Eye position was sampled at 1000 Hz and saccades prior to critical fixations were detected online with a nine-sample

saccade detection model using a velocity trigger of 50°/s. Viewing was binocular, but only the right eye was tracked. Stimulus presentation and response recording were controlled via Experiment Builder (SR Research, Canada).

Materials

Each participant completed both a text reading and a scene viewing condition. In the reading condition, participants read 25 pages of text. Each page presented a story adapted from Aesop's fables. In total, the 25 fables comprised 2923 words. Each fable and display page was composed of 100 to 133 (mean 116.9) words, distributed across 11 to 13 left-justified lines of text, with line spacing of 1.35° (42 pixels). Maximum line length was 50 characters, and one character (14 pixels) subtended 0.45 degrees of visual angle horizontally. Word lengths ranged between 1 and 14 characters, with a mode of 3 and a mean of 4 characters. The text was presented in black on a white background.

In the scene viewing condition, participants viewed 40 unique full-color 800 × 600 pixel photographs of real-world scenes (20 indoor, 20 outdoor) from a variety of scene categories. Each scene subtended a visual angle of 25.78° × 19.48°.

Procedure

Participants were instructed to remember the texts for a subsequent comprehension test and to memorize the scenes for a subsequent memory test, neither of which was administered. Participants were told to ignore any occasional flicker they might notice.

Stimulus onset delay was implemented using a saccade-contingent display change technique, where every sixth saccade was manipulated as follows (Figure 3, supplementary Movie 1): The stimulus was erased from the CRT and replaced by a mask during the saccade, when visual transients were suppressed (Ross, Morrone, Goldberg, & Burr, 2001). When the eyes landed in the critical fixation following this saccade, the stimulus was no longer visible. Following the predetermined delay, the stimulus reappeared. Participants were presented

delays of 0 (control), 300, 400, 600, 800 ms, or infinite. In the 0-ms delay control condition, the stimulus (text or scene) was replaced with itself so that phenomenally it was continuously present but the computer code generating changes in the other conditions was controlled. In the infinite delay condition the stimulus only reappeared when participants moved their eyes to end the critical fixation (cf., Yang & McConkie, 2001). To avoid phosphor persistence, a color noise mask was presented during the delay.

In the text reading condition, the setup differed from previous text onset delay studies (Morrison, 1984; Rayner & Pollatsek, 1981) in a number of ways. First, we used paragraph-long, multi-line sentences as opposed to single sentences. In addition, in the present study we used comparatively long delays. The shortest delay in our study (300 ms) coincides with the longest delay condition in previous studies. Second, in previous studies the text was delayed during each fixation, whereas in the present study the delay only happened every sixth fixation. Third, the stimulus onset was delayed with the same full-display noise mask in the scene viewing and the text reading conditions. As a result, the mask obscured all lines of text in the text reading condition. In comparison, in previous studies the visual mask (an interlaced square wave grating) covered the entire line of text or a number of characters around the current point of gaze only (Morrison, 1984; Rayner & Pollatsek, 1981). Display changes administered during eye fixations can lead to saccadic inhibition (Reingold & Stampe, 2002, 2004), and previous research has shown that large sudden onsets (i.e., display changes) produced stronger saccadic inhibition than small flickers (Reingold & Stampe, 2003). In the present experiment, a full-screen mask was used. When the mask is removed to reveal the stimulus at the end of the delay, saccadic inhibition is likely to be stronger than when using a smaller mask. However, the larger mask allowed us to compare scene viewing and reading under similar onset delay conditions. Also, saccadic inhibition would not be expected in the infinite delay condition since all display changes in that condition took place

during saccades.

An experimental trial took place as follows. First, calibration was checked. Then, a fixation cross was presented at a top left location on the screen. Once the participant had fixated the cross, the trial was initiated. In the scene viewing condition, each scene was presented for 40 saccades, allowing the implementation of one instance of each delay condition in each trial. In the reading condition, each text was presented until the subject pressed the spacebar on the keyboard to signify that they had finished reading. Within each scene or text, delay values were chosen pseudo-randomly for each critical fixation. Presentation order of text reading and scene viewing blocks was counterbalanced across subjects and the presentation order of items was randomized within a given block.

Insert Figure 3 about here

Gaze Data Analysis

Saccades were defined with a 50°/sec velocity threshold using a nine-sample saccade detection model. Raw data were converted into a fixation sequence matrix using SR Research Data Viewer.

Results

Behavioral Data

The scene onset delay data from this experiment were previously used for a CRISP simulation study reported in Nuthmann et al. (2010). The present analyses focus on the text reading data and on comparison with the scene viewing data.

Trial duration. The average trial duration was 14.1 sec for scene viewing and 31.9 sec for text reading. Critical fixations that began before the stimulus was completely erased from the CRT and in which the participant blinked were removed from analysis. For the

scene viewing data, there were on average 26.6 (out of 40) valid data points per participant and delay condition. In the text reading condition, where trial duration was determined by participants' reading behavior, 25 texts provided an average of 64.3 valid critical fixations per participant and delay condition. Thus, statistical power was greater in the text reading than in the scene viewing conditions.

Total number of fixations during the delay. The delays tested in the present study varied between 300 and 800 ms. As a first global manipulation check, we analyzed the number of fixations that participants made, on average, when the stimulus was delayed. This analysis excluded the infinite delay condition since there was always only one fixation. For the other delay conditions, Figure 4 displays the results in a stacked bar graph. Each delay condition on the x-axis is represented by two bars. The left bar (T) represents the text reading data and the right bar (S) the scene viewing data. Each stack displays the relative frequencies of the eyes making exactly 1, 2, 3, 4, or ≥ 5 fixations during stimulus absence. When there was only one fixation during stimulus onset delay, the eyes didn't move before the stimulus reappeared. The probability of a single fixation decreased as the delays got longer. For delays longer than 300 ms, there was an effect of task: Participants were more likely to wait for the stimulus to return when reading a text as opposed to viewing a scene. This was true even though the average fixation duration was shorter in reading than in scene viewing (0-ms delay control condition in Table 2, Figure 5). In contrast, when there was more than one fixation during stimulus delay, the critical fixation was terminated while the mask was present, and one or more additional fixations were made before the stimulus reappeared. Generally, as delays got longer participants tended to make more additional fixations before the stimulus reappeared.

Insert Figure 4 about here

Mean fixation durations. The remaining analyses considered the duration of critical fixations. First, the data from the control condition confirm that fixation durations in reading have a shorter average duration than in scene perception (Table 2), and the range of fixation durations is smaller (Figure 5), a common finding (Henderson & Hollingworth, 1998). With regard to the stimulus onset delay manipulations, if all fixations were directly controlled by the text or scene, then the observed critical fixation durations should increase in proportion to the duration of the stimulus onset delay (Rayner & Pollatsek, 1981; see Figure 3 in Yang, 2009, for visualization). That is, critical fixation durations are predicted to be the sum of the mean fixation duration (Table 2: reading 184 ms, scene viewing 262 ms) and the stimulus onset delay. This relationship is formalized by the following regression equation:

$$\text{critical fixation duration} = \text{mean fixation duration} + 1 * \text{delay duration} \quad (2)$$

If fixation durations perfectly reflect the availability of useful visual information, then the slope of the empirical fixation duration function should approach 1.0. In the present study, average critical fixation durations increased with onset delay in both text reading and scene viewing (Table 2). For statistical analyses, one-way repeated measures ANOVAs with delay (0, 300, 400, 600, 800 ms) as the factor were performed on data from each task. In both tasks, the increase in fixation duration across delays was significant [text reading: $F(4,11) = 90.53$, $p < .001$; scene viewing: $F(4,11) = 22.44$, $p < .001$]. However, the slope of the relating fixation duration to delay was not as steep as predicted by the linear increase hypothesis (text reading data: slope = 0.22, y -intercept = 312 ms; scene viewing data: slope = 0.24, y -intercept = 331 ms). This is not surprising given that, as discussed earlier, a certain percentage of critical fixations were terminated during the delay (Figure 4).

Insert Table 2 about here

Fixation duration distributions. Logically, the increase in fixation duration in the delay conditions can result from a certain amount of lengthening of all critical fixations, or from an even greater lengthening for only some subset of the critical fixations. Previous analyses have shown that two groups of fixations are found in this paradigm, one influenced by delay and the other unaffected by delay (Henderson & Pierce, 2008; Henderson & Smith, 2009; Morrison, 1984; Nuthmann, et al., 2010). To investigate this issue in the present study, Figure 5 shows the distributions of fixation durations as a function of delay condition and task. Each panel compares the text and scene data for a given delay condition or the 0-ms delay control condition. Vertical broken lines mark the duration of the delay as a reference point. Figure 6 provides an additional visualization where all distributions are superimposed, which allows for a more direct comparison of the different delay conditions with the 0-ms delay control condition.

With respect to the fixation duration distributions, there are two initial comparisons to make. First, how do the onset delay conditions in each task compare to their respective 0-ms delay control condition, and second, how do the onset delay conditions compare across the text reading and scene viewing tasks. As can be seen in Figure 5, for the non-infinite delay conditions, fixation duration distributions appeared to be bimodal in both the text reading and the scene viewing tasks. Two-term Gaussian models [$a_1 * \exp(-((x-b_1)/c_1)^2) + a_2 * \exp(-((x-b_2)/c_2)^2)$] were fit to the data from these delay conditions, separately for each task and delay (Figure 5). One-term Gaussian models were fit to the fixation duration distributions from the 0-ms delay control condition and the infinite delay condition. Each term of a fitted Gaussian distribution is described by a mean and standard deviation. The mean denotes the location of the peak, and the standard deviation describes how much the values spread around the mean. In reading, it appears that the first peaks of the 300 to 800-ms delay conditions

were somewhat shifted towards longer fixation durations when compared to the 0-ms delay normal reading control condition (Figure 6a). Also, the standard deviation was larger than in the control condition, and it somewhat increased with onset delay. This rightward shift in fixation duration distributions was also observed for the scene viewing data, though there was no clear pattern concerning the standard deviations. With regard to the second modes of the distributions, they appeared more peaked (smaller standard deviations) in reading compared to scene viewing.

Insert Figure 5 about here

Insert Figure 6 about here

For a given stimulus onset delay, fixations with durations longer than the delay reflect those for which the eyes waited until the stimulus returned before moving. Fixations with durations shorter than the delay represent those for which the eyes moved before the stimulus returned. For analysis, one could simply assign each fixation to one or the other population. However, it appears that the composite bimodal distributions (Figure 5) represent a mixture of two more-or-less overlapping distributions. Therefore, the modes of the fitted two-term Gaussian distributions were analyzed rather than the fixation duration means for the two populations (cf., Henderson & Pierce, 2008; Henderson & Smith, 2009). For each task, we then performed regression analyses over the first and second modes of the fitted distributions (Figure 7). For the second (or late) distribution modes, these analyses demonstrate that fixation durations increased linearly and in a one-to-one relationship with delay, in both reading and scene viewing: text reading data, slope = 0.99, y-intercept = 185 ms; scene

viewing data, slope = 1.07, y -intercept = 183 ms. For the reading data, the intercept matches the average fixation duration in reading (185 ms vs. 184 ms). For the scene viewing data, the intercept is lower than predicted by the linear increase hypothesis (183 ms vs. 262 ms). However, these intercepts should be interpreted with caution; according to our understanding of saccadic inhibition and its implementation in the CRISP model, they do not directly translate to average fixation durations (see below).

A second pair of regression analyses was performed over the first (or early) modes of the fitted distributions (Figure 7). The results showed that fixation durations remained constant across delay: text reading data, slope = 0.06, y -intercept = 249 ms; scene viewing data, slope = 0.04, y -intercept = 250 ms. The regression slopes did not significantly differ from 0 ($p > .05$). Overall, this pattern of data for scene viewing closely replicates the results from past scene viewing studies using the same paradigm (Henderson & Pierce, 2008; Henderson & Smith, 2009). The qualitative similarity of the pattern for reading in the same paradigm is novel and interesting.

Insert Figure 7 about here

Saccadic inhibition and infinite delay. Apparent in Figure 5 is a dip in the fixation duration distribution at each non-infinite text or scene onset delay. This dip likely reflects saccadic inhibition induced by the reappearance of the text or scene after the delay (Reingold & Stampe, 2002, 2004). To what extent could saccadic inhibition be responsible for the increased fixation durations in the stimulus onset delay conditions? Maximum saccadic inhibition typically occurs 100 ms following the onset of a visual change, and the inhibition period is followed by a recovery period with increased saccadic activity (Reingold & Stampe, 2002, 2004). The saccadic inhibition signature depends on the characteristics of the baseline

fixation duration distribution obtained in the absence of a visual change, and is strongest when the onset time of the visual change equals median baseline saccadic reaction time minus 100 ms (Reingold & Stampe, 2002, their Figure 1). This logic can be applied to the present data by using the mean fixation durations in the 0-ms delay baseline conditions as reference (Table 2). Accordingly, the stimulus onset delays that are expected to produce maximum saccadic inhibition are 84 ms for text reading ($184 - 100 = 84$ ms) and 162 ms for scene viewing ($262 - 100 = 162$ ms). In the present experiment, the reinstatement of the stimulus took place between 300 and 800 ms following the onset of the critical fixation. Those delay durations should be too long to produce major dips in fixation duration distributions due to saccadic inhibition, because they hit the baseline distributions toward the tails or even beyond the tails (see 0-ms delay baseline distributions in Figure 5). This is particularly striking in the reading task where 93% of valid fixations in the normal reading control condition were shorter than 300 ms. Still, we observe clear bimodality in fixation duration distributions for the non-infinite onset delay conditions (Figure 5, Figure 6) because a considerable number of directly controlled fixations are lengthened beyond the duration of the delay. If they were not delayed, there would be no fixations from which to create a saccadic inhibition dip. Therefore, even though saccadic inhibition appears to lengthen fixation durations, it cannot alone account for the longer fixation durations observed in the stimulus onset delay conditions (cf., Henderson & Pierce, 2008; Henderson & Smith, 2009).

The experiment included an infinite delay condition to test directly whether effects of the delay indicating direct control of fixation durations were contaminated by saccadic inhibition related to motion transients during fixations. In the case of an infinite delay, both the initial critical stimulus disappearance and the subsequent stimulus reappearance took place during a saccade. Consequently, there were no motion transients during fixations to produce saccadic inhibition. Conceptually, the infinite delay differed from the other delay

conditions in that holding fixation until the stimulus returned proved to be an unsuccessful strategy. However, this was not recognizable to the participants as infinite delays were randomly intermixed with the other delay conditions. Indeed, until 800 ms into the critical fixation, the infinite delay was indistinguishable from the longest delay condition.

The data from the infinite delay condition are informative in several ways. First, these data provide further evidence for direct control of fixation durations. In both tasks, on average critical fixations lasted longer in the infinite delay compared to the 0-ms delay control condition (Table 2), and the corresponding distributions were shifted towards longer fixation durations (Figure 5, Figure 6). Second, there were differences in the adjustment of fixation durations in the two tasks. The average fixation duration in reading was found to be shorter and the corresponding distribution much more peaked than in scene viewing (Table 2, Figure 5). Yet in the infinite delay condition, participants were more inclined to prolong fixations in reading than in scene viewing, as is evident from a broader fixation duration distribution for reading compared to scene viewing (Figure 5). Third, for each task the initial part of the infinite delay fixation duration distribution corresponded quite well with the first modes in the non-infinite delay conditions, as can be seen from the superimposed distributions in Figure 6. All of these results suggest that the fixation duration distribution obtained for the infinite delay is the more appropriate baseline distribution for evaluating effects of saccadic inhibition on fixation durations than is the 0-ms delay condition. In the infinite delay condition, 63.3% (reading) and 55.1% (scene viewing) of critical fixations were longer than 300 ms, leaving enough fixations to create a saccadic inhibition dip during the 300-ms delay.

Simulated Data

The data from the scene onset delay condition in the experiment have been previously simulated with the CRISP model (Nuthmann, et al., 2010). The simulated data qualitatively

reproduced the two populations of fixation durations observed in the empirical data. In particular, there was good qualitative agreement between simulated and empirical fixation duration distributions (Nuthmann, et al., 2010). To test the generality hypothesis outlined above, in the present simulations the very same model architecture was applied to the data from the text onset delay condition. Thus, model generalizability was analyzed in the restricted sense of parameter changes. Generally, this is a much more stringent test than adding new parameters to the model (Nuthmann & Engbert, 2009).

Model adjustments. CRISP simulations of fixation duration data from the stimulus onset delay paradigm allow testing a small set of simple rules for the modulation of saccade timing and saccade programming by visual-cognitive processing. Specifically, in CRISP current processing demands modulate the random walk's transition rate, and processing difficulty can lead to saccade cancellation (Nuthmann, et al., 2010, for the scene onset delay data). The details of these assumptions will be discussed next, now referring to the text onset delay data.

First, the model assumes that difficulties in moment-by-moment visual and cognitive processing lead to adjustments in the random walk's transition rate. The general principle was introduced above, including visualizations in Figure 2. In the text onset delay simulations it was adapted as follows: When the text is removed from view during the onset delay, the mean random walk transition rate r_I is considerably reduced. The implementation takes an eye-brain lag of 50 ms into account (e.g., Sereno, Rayner, & Posner, 1998), so 50 ms following text offset, the mean transition rate is reduced from r_I to r_0 . Thus, the random walk process is slowed down, which delays the initiation of the next saccade program. Fifty milliseconds after the text reappears, the rate recovers to its default value r_I .

Saccade cancellation provides a second mechanism contributing to prolonged fixation durations in the stimulus onset delay paradigm. The underlying rationale is that removing the

text from view interrupts the preparation of eye movements. If a labile saccade program is active when the text disappears, it is subject to stochastic cancellation. This processing-related saccade cancellation mechanism is also subject to the 50-ms eye-to-brain lag. A second cancellation mechanism was implemented as response to the text reappearance during fixation (see below). A visualization of the corresponding implementation in the scene onset delay paradigm is provided in Nuthmann et al. (2010, Figure 7).

In sum, the model comprises parameters related to saccade timing (t_{sac} , N , r_0/ r_1) and saccade programming (τ_{lab} , τ_{nlab} , τ_{ex}), including two probabilities of saccade cancellation ($p1_{canc}$, $p2_{canc}$), all of which are summarized in Table 3. For simulation of the text onset delay data, the mean duration of saccade execution (τ_{ex}) was fixed at 40 ms (the value was estimated from the saccade durations in the experiment). For all other parameters, best-fitting values were determined with a genetic algorithm optimization technique (Sivanandam & Deepa, 2007). The genetic algorithm minimized a goodness-of-fit measure, which quantified how much the simulated fixation duration distribution and average fixation duration deviated from the experimentally observed data. The ranges for the parameter values were informed by findings from basic oculomotor research, which ensured their psychological and/or neurophysiological plausibility. The details of the general fitting procedure are provided in Appendix B in Nuthmann et al. (2010). The corresponding best-fitting parameter values are listed in Table 3. To facilitate comparisons, parameter estimates for the scene onset delay data, taken from Nuthmann et al. (2010), are also presented.²

Insert Table 3 about here

Text onset delay data. For the most part, fixation duration analyses in the reading and scene perception literatures occur at the level of the means. Analyses of the empirical

stimulus onset delay data convincingly demonstrate that the changes in mean critical fixation durations reflect distinct patterns at the level of the underlying distributions. Thus, the primary goal of the simulations was to go beyond the mean and reproduce the distributions for the critical fixations. The behavioral data show a number of key features, which must be captured by any simulated data. First, the simulations must replicate the relatively small range of fixation durations observed in normal reading. Second, they must reproduce lengthened critical fixation durations in the infinite delay condition, including a shift of the modal portion of the distribution towards longer fixation durations, and an increased tail. Third, for all other delay conditions the simulations must reproduce the typical bimodal fixation duration distributions, including an accurate proportion of fixations that were prolonged beyond the duration of the delay. The data from the different delay conditions were simulated and fit jointly in an implementation that closely mirrored the sequence of events in the experiment. Specifically, simulated text onset delays took place every sixth saccade, and delay values were chosen pseudo-randomly for each critical fixation. Simulated sequences of fixation durations were obtained from 12 statistical subjects and 25 arbitrary texts per subject, using the best-fitting values for model parameters (Table 3).

Figure 8 plots the resulting fixation duration distributions for the critical fixations. Each panel compares the simulated and empirical data for a given text onset delay condition or the 0-ms delay control condition. Overall, the simulations with the CRISP model captured the fixation duration distributions well. The fit was qualitatively accurate and quantitatively satisfactory. First of all, the simulations replicated the relatively small range of fixation durations observed in normal reading. In addition, for all non-infinite delay conditions, the simulations qualitatively reproduced the typical bimodal fixation duration distributions. The infinite delay condition proved to be an important boundary condition as it allowed us to determine the extent to which the results were contaminated by saccadic inhibition.

Compared to the control data, the empirical fixation duration distribution for the infinite delay was skewed toward longer fixation durations, indicating that fixation durations are under direct control of the current visual input. This signature was well captured by the simulated data. Having eliminated saccadic inhibition in the infinite delay, unimodal distributions were observed and reproduced. According to the model architecture, simulated fixation durations were lengthened due to an adjustment in the random walk transition rate of the timing signal and processing-related saccade cancellation. In the case of an infinite delay the text reappeared during a saccade, while it was restored during a fixation in the other delay conditions. In the CRISP framework, a significant proportion of currently labile saccade programs is cancelled in response to a visual display change during fixation (Nuthmann, et al., 2010). In the stimulus onset delay simulations this had the effect of prolonging the latency from stimulus reinstatement. Model simulations produced a unimodal distribution for the infinite delay and bimodal distributions for all other delays, suggesting that this additional saccade cancellation mechanism indeed contributed to the saccadic inhibition dip observed in these distributions. The fit was less satisfactory in the 300-ms delay condition: The second mode of the distribution, reflecting the recovery from saccadic inhibition, was less peaked in the simulated data than in the empirical data.

Insert Figure 8 about here

We conclude this section with a comment on how task-specific influences were realized in the model simulations. To account for global characteristics of fixation durations in reading and scene viewing, we assumed that participants implement different global parameter settings when reading texts as opposed to viewing scenes. The reading data were fit independently of the scene viewing data (see also footnote 1), and we allowed all free

model parameters to vary across tasks. As a result, both saccade timing and saccade programming parameters differed across tasks (Table 3). Specifically, for the reading task a lower mean value for the random timer (t_{sac}), accompanied by smaller variance (larger N), was obtained. Shorter saccade latency, defined as the sum of labile and non-labile stages of saccade programming, was estimated for reading compared to scene viewing. Generally, a profound psychological interpretation of these differences appears to be out of place as we did not test specific hypotheses, and did not fit the data from the two tasks simultaneously.

General Discussion

The goal of the present study was to compare the control of fixation durations in reading and scene viewing. The stimulus onset delay paradigm (e.g., Henderson & Pierce, 2008; Rayner & Pollatsek, 1981) was used to investigate whether and how fixation durations are controlled ‘online’ by the current visual input (a text or a scene). Key empirical findings were simulated with the CRISP model of fixation durations (Nuthmann, et al., 2010), with task-specific influences realized by different parameter settings. The results suggest that global characteristics of fixation durations in scene viewing and reading can be explained by a common control system.

Generalizability of Eye-Movement Models across Tasks

A good computational model of cognition must fulfill the criterion of generalizability (e.g., Pitt, et al., 2002). Generalizability refers to the model’s ability to account for more than one effect in one particular task. With regard to reading models, it has been suggested that models of eye-movement control in reading must have the potential for generalization to non-reading tasks (Engbert, et al., 2005; Nuthmann & Engbert, 2009). It appears that the issue requires differentiated views with regard to different non-reading tasks. Of all contemporary reading models, the SWIFT model was designed as a general model of eye-movement

control, where reading is looked upon as a case study (Engbert, et al., 2005). In a recent simulation study (Nuthmann & Engbert, 2009), the model's generality beyond reading was successfully tested with a *z*-string scanning paradigm, where all letters of the text are replaced by the letter *z* (preserving spaces, punctuation and case sensitivity). *Z*-string scanning approximates reading without lexical processing and shares the visuomotor requirements with reading. In comparison, the similarities between reading and non-reading tasks like scene perception and visual search, where eye movements operate in two dimensions, are less obvious.

When eye-movements are compared across reading, scene perception, and search, fixation durations and saccade lengths in reading do not correlate well with those measures in scene perception and search (Andrews & Coppola, 1999; Rayner, et al., 2007). This has been taken as evidence that the mechanisms that guide the eyes during reading are specific to the reading task (Rayner, 2009a; Rayner, et al., 2007). In a recent review, Rayner (2009a) argued that it is “somewhat hazardous to generalize across these tasks in terms of eye movement behaviour” (p. 1459). There is no doubt that reading is unique; reading and scene viewing both engage vision, but reading also engages language. With the present work we asked a more nuanced question: At what level of a model's architecture does reading differ from other tasks? The approach taken here is that at the most basic level of saccade timing and programming, reading and scene viewing are not fundamentally different. Irrespective of the task, eye movements are made to compensate for the lack of high resolution vision outside of the fovea, the same neural circuitry is involved in controlling these movements, and how the cognitive system interacts with the oculomotor system is subject to the restrictions that arise from the operation of the oculomotor system.

Stimulus Onset Delay Experiment: Mixed Control of Fixation Durations in Reading and Scene Viewing

We investigated the control of fixation durations in reading and scene viewing by using the stimulus onset delay paradigm. This paradigm is particularly suitable for task comparisons as it can be applied to visual stimuli of all kinds, and it selectively manipulates global stimulus processing difficulty. In a within-subject design, participants read texts and viewed pictures of real-world scenes presented on a computer screen, with the same implementation of stimulus onset delays in both tasks. A saccade-contingent display change technique was used to present a full-screen visual mask at the beginning of specific critical fixations, which delayed the onset of the text or scene. The duration of the delay varied between 300 and 800 ms. In an additional infinite delay condition the stimulus only reappeared when participants moved their eyes to end the critical fixation. The main question was how the duration of the critical fixation would be affected by the onset delay.

The results showed that the distribution of fixation durations changed as a function of delay. Irrespective of task, two populations of fixation durations were observed. One population of fixations increased in duration as delay increased, suggesting that they were under the direct control of the current stimulus. A second population of fixation durations was relatively constant across delay, suggesting that they were insensitive to the current visual input. This qualitative pattern was observed for both reading and scene viewing. However, additional task-specific quantitative differences in the adjustment of fixation durations were found. Any interpretation of task effects must take the differences in baseline fixation duration distributions into account (see Yang, 2009; Yang & McConkie, 2001, for an analysis of saccade hazard levels). These baseline distributions were provided by data from the 0-ms delay control condition. Fixation duration was found to be less variable for reading, with fewer long fixations than in scene viewing (Figure 5), which accords with previous findings (Henderson & Hollingworth, 1998). In the experiment, the minimum stimulus onset delay was 300 ms. In the baseline distributions, only 7% (reading) or 29% (scene viewing) of

valid fixations were longer than 300 ms. Analyses of the number of fixations made during the delay (Figure 4), and the duration of critical fixations in the non-infinite and infinite delay conditions (Table 2, Figure 5, Figure 6), all demonstrated that participants were more likely to wait for the stimulus to return when reading a text as opposed to viewing a scene. Apparently, delaying the stimulus was more disruptive in reading than in scene viewing, which suggests that the availability of useful visual information during fixation is of even greater importance in reading.

For each non-infinite text or scene onset delay there was a dip in the fixation duration distribution, reflecting saccadic inhibition induced by the reappearance of the text or scene after the delay (Reingold & Stampe, 2002, 2004). However, saccadic inhibition cannot be the sole account for increased fixation durations observed in the stimulus onset delay conditions (cf., Henderson & Pierce, 2008; Henderson & Smith, 2009). The most direct evidence supporting this conclusion was obtained from the infinite delay condition, where fixation durations are not affected by saccadic inhibition. When encountering an (unpredictable) infinite delay, participants considerably prolonged their fixation durations compared to the 0-ms delay control condition even though no saccadic inhibition was produced (Table 2, Figure 5, Figure 6). With regard to saccadic inhibition, the present results are in line with a study by Yang (2009) that attempted to disentangle the effects of processing difficulty and visually-induced saccadic inhibition in experiments utilizing the gaze-contingent display change technique. Using a dual text-change paradigm, he found that a combination of both factors is responsible for the changes in fixation duration.

Modeling of Stimulus Onset Delay Data with the CRISP Model

The data from the scene onset delay condition in the experiment have been previously used to validate the CRISP model of fixation durations (Nuthmann, et al., 2010). In the present work, the model's generality was tested by applying its architecture to the text onset

delay data. The general approach to modeling these data was to realize global task-specific influences by different model parameter settings.

In brief, in the CRISP model there are two separate though strongly interacting time lines that combine to produce fixation durations in the model: random walk timing signals and saccade programming involving labile and non-labile stages (Nuthmann, et al., 2010). The random walk timing signal accumulates toward a threshold and initiates a new saccade program once that threshold is reached. Saccades can be programmed in parallel, and a later saccade program can cancel an earlier one if it is still in its labile stage of programming. The visual-cognitive processing module directly controls both the saccade timer and the saccade-programming module (Figure 1). On the saccade programming time line, processing difficulty can lead to the cancellation of saccade programs that are currently in the making. The random walk saccade timer is controlled via modulation of its transition rate. Thus, visual-cognitive processing demands continuously and directly adjust saccade timing, although saccade timing itself is not coupled to certain stages of cognitive processing. Due to the conceptualization of saccade timing and programming in the model, the model simulations reproduced the mixed control of fixation durations observed in the stimulus onset delay paradigm, tested on both text reading and scene viewing. Roughly speaking, whether or not delaying stimulus information influences the critical fixation duration depends on how far along saccade planning has proceeded (cf., Morrison, 1984). Sometimes it will be too late to allow any influence. These will most likely be instances where the autonomous saccade timer initiated the saccade program prior to the onset of the current fixation, resulting in comparatively short fixations that end during the delay (see Nuthmann, et al., 2010, for in-depth discussion). At other times inhibitory signals from the visual-cognitive processing module will delay the start of a new saccade program, eventually prolonging fixation durations. In addition, the empirical data showed signs of saccadic inhibition due to visible

changes in the display (Reingold & Stampe, 2002, 2004). In the model, saccadic inhibition was accounted for by stochastic cancellation of saccade programs.

Direct Control of Fixation Durations in Reading

Thirty years have passed since the stimulus onset delay paradigm was first used to investigate the direct control of fixation durations in reading (Morrison, 1984; Rayner & Pollatsek, 1981). Since then, important advances have been made in understanding eye guidance in both reading and scene viewing. It is now generally accepted that the time the eyes spend on a word in reading is modulated by a range of low-level visuomotor and high-level linguistic factors (see Rayner, 1998; 2009a, for reviews). However, the debate concerning the direct control of fixation durations is still an ongoing one. In the case of reading, the direct lexical control hypothesis states that for the majority of reading fixations, the processing of the lexical properties of the fixated word influences the timing of the saccade terminating that fixation (e.g., Rayner & Pollatsek, 1981; Rayner, et al., 1996). The current debate is mostly concerned with the time course of lexical influences and the proportion of reading fixations that are impacted by lexical variables (see Reingold, Yang, & Rayner, 2010; Staub, White, Drieghe, Hollway, & Rayner, 2010, for discussion).

With the present work, we took a step back and revisited the classic text onset delay paradigm (Morrison, 1984; Rayner & Pollatsek, 1981). Given the empirical results and numerical simulations obtained here for the text reading data, the question arises whether current models of eye-movement control in reading can principally explain the signature findings from the text onset delay paradigm.

Interestingly, the first instantiation of the E-Z Reader model (Reichle, et al., 1998) set out to be a quantification and elaboration of Morrison's (1984) qualitative model and its subsequent modifications. Morrison's (1984) explanation for the data pattern observed in his text onset delay experiments was based on a consequence of parallel programming of

saccades: “If saccades during reading can be programmed in parallel, then some saccades will not appear to be programmed in response to the immediately preceding fixation, because they were programmed or initiated before the information had been processed or even during the prior fixation” (p. 678). E-Z Reader departs from Morrison’s (1984) model in that eye-movement programming is decoupled from shifts of covert attention (Reichle, et al., 1998). In both models, saccades can be programmed in parallel, and a later saccade program can cancel an earlier one. However, in the E-Z Reader model new saccade programs are typically not initiated before the start of the fixation they terminate. Instead, E-Z Reader implements design principles reflecting strong direct control of fixation durations. In the model, saccade programming is time locked to word processing. Specifically, a new saccade program is initiated when the word processing system has completed an initial stage of processing, referred to as L_1 (originally termed a *familiarity check*), on the fixated word. Therefore, in most cases saccade programs terminating the current fixation are initiated only after the start of that fixation. The completion of a later stage of word processing (L_2 , originally termed *lexical access*) on one word causes attention to shift to the next word so that L_1 can begin on that word.

One way E-Z Reader could potentially account for the increased fixation durations observed in the text onset delay paradigm is to assume that processing difficulties induced by the text onset delay somehow extend the duration of the early lexical processing stage L_1 . With regard to the fixations that ended while the mask was still present, one could argue that subjects sometimes moved their eyes during the mask because they could not always keep them fixated for lengthy times (Rayner & Pollatsek, 1981). To account for these fixations, it could be assumed that saccade initiation is proportionally delayed by the masking of text content until an oculomotor deadline is reached, at which time a saccade is automatically triggered (Henderson & Ferreira, 1990). It is conceivable that these mechanisms could

reproduce an increase in fixation durations and/or corresponding shifts in fixation duration distributions. However, they would not reproduce an apparent key feature of the text onset delay signature, that is the bimodality in fixation duration distributions with early modes that are constant across delays, and second modes that increase linearly with delay (Figure 5, Figure 7). In the present CRISP simulations, processing-related saccade cancellation was implemented as one of two mechanisms contributing to prolonged fixation durations in the stimulus onset delay paradigm. At the time point of stimulus disappearance, if there was a labile saccade program active, it was subject to stochastic cancellation. It appears that such a cancellation mechanism is less compatible with the E-Z Reader architecture. Conceptually, it appears less sensible to have L_1 run to completion (while the text is blocked from view during the delay) to initiate a new saccade program only to cancel it right away. In addition, when the text disappears at the beginning of a critical fixation, often there will be no active labile saccade program that can be cancelled.

In sum, it appears that E-Z Reader implements a strong version of direct control of fixation durations, and this conceptualization might be too constrained to reproduce the mixed control of fixation durations observed in the text onset delay paradigm. In contrast, the signature finding from the stimulus onset delay paradigm can be accounted for by a model advocating a weaker coupling between eye-movement programming and processing of the currently fixated stimulus. In particular, we interpret the results of the present study as support for autonomous timing and temporally overlapping saccade programming as implemented in CRISP (Nuthmann, et al., 2010) and SWIFT (Engbert, et al., 2005; Schad & Engbert, 2012 this issue).

CRISP and SWIFT Models

What is the relationship between the CRISP and SWIFT models? Both models share the core assumption that an autonomous saccade timer initiates saccade programs after

random time intervals, while these saccade initiation intervals can be adjusted by ongoing processing demands. However, the models differ in their conceptualization and implementation of this basic idea. In SWIFT 2 (Engbert, et al., 2005), autonomous timing signals are drawn from a gamma distribution, while the sampled time interval can be prolonged by a high activation value of the currently fixated word. In contrast, in CRISP (Nuthmann, et al., 2010) timing signals are implemented as discrete-state continuous-time random walks. An obvious advantage of such a conceptualization is the way processing difficulties modulate the timer, which can be implemented as a *continuous* process (Nuthmann, et al., 2010). More recently, the random-walk implementation of the autonomous timer has been adopted and extended by the ICAT model (Trukenbrod & Engbert, submitted) and subsequently by a third iteration of the SWIFT model (SWIFT 3: Schad & Engbert, 2012 this issue). In these models, random walks control not only saccade initiation intervals (as in CRISP) but also the various stages of saccade programming. Furthermore, CRISP and SWIFT differ quite substantially in the exact mechanisms by which processing difficulties modulate the random timer.

Allocation of Visual Attention in Reading and Scene Viewing

The present study investigated a specific aspect of attention allocation in reading and scene viewing, i.e., the mechanisms that control *when* attention or the eyes move. The remainder of the General Discussion concerns the mechanisms that control *where* attention or the eyes move. Specifically, we discuss whether modeling principles that were developed to account for the allocation of visual attention in reading could principally be applied to scene viewing. In E-Z Reader, visual attention is allocated serially from one word to the next and serial lexical processing is invoked (word n is identified, then word $n+1$, etc.) (Reichle, et al., 1998; Reichle, et al., 2003; Reichle, Warren, et al., 2009). Thus, word $n+1$ is always the default saccade target. In contrast, the SWIFT model envisages a gradient of attention within

which a degree of parallel lexical processing can occur so that more than one word can be processed at a time (Engbert, et al., 2005). Saccade target selection is probabilistic in that the word with the highest activation in a dynamically changing field of activations has the highest probability of being selected as the target for the next saccade. The nature of attention allocation is currently an issue of much contention in the literature on eye-guidance in reading (Reichle, Liversedge, et al., 2009). The issue is not well explored in object and scene perception, but it bears much less potential for contention. When viewing a scene as opposed to reading a text, there is no inherent spatial order in which objects need to be processed, which may hinder the extrapolation of a E-Z Reader like sequential attention shift architecture to scene viewing (see De Graef & Germeys, 2003, for discussion). In comparison, the SWIFT architecture is more compatible with selection from a set of potential saccade targets.

At the same time, while it is clear that scene-level features can be processed across the visual field during scene viewing, it has yet to be shown that objects are in fact processed in parallel in scene perception. Instead, the evidence from several paradigms suggests that focal attention is needed to recognize and encode objects into memory. For example, there is considerable evidence from the change blindness literature that changes to objects are not noticed unless the changing object is focally attended (Henderson & Hollingworth, 1999, 2003; Rensink, O'Regan, & Clark, 1997; Simons, 2000). If objects were processed in parallel across the scene, one would expect that such changes would be easily noticed. Similarly, semantically odd or emotional objects do not appear to immediately “pop out” in a scene as one might expect if objects were processed in parallel (see Underwood, 2009, for a review). In normal scene viewing, focal attention and fixation tend to be tightly coupled (e.g., Hollingworth & Henderson, 2002). From this perspective, the mechanism for determining

where the eyes move next in scenes may not be as incompatible with the serial assumptions of E-Z Reader as would first appear. These issues await further investigation.

Outlook

The stimulus onset delay paradigm provides an existence proof that fixation durations can be modulated in real time by the stimulus available in a fixation. In the reading literature it is widely acknowledged that properties of the text, including local word properties as well as syntactic and discourse factors, exert an immediate influence on fixation durations (Rayner, 1998). In analogy, further research is warranted to determine whether fixation durations in scene viewing are also influenced by more subtle scene properties. This could open the door to using fixation durations as a moment-to-moment online index of attention and ongoing perceptual and cognitive processes during scene viewing, as has been done to great benefit in the reading literature. With regard to the issue of model generality, further empirical and computational research is required to test control principles of fixation durations within and across tasks.

References

- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39(17), 2947-2953.
- Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19, 1967-1983.
- Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology*, 62(1), 1-14.
- De Graef, P., & Germeys, F. (2003). Reading the scene: Application of E-Z Reader to object and scene perception. *Behavioral and Brain Sciences*, 26(4), 479-480.
- Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5), 621-636.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777-813.
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22(4), 661-721.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498-504.
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4), 219-222.
- Henderson, J. M. (2011). Eye movements and scene perception. In I. D. Gilchrist, S. P. Liversedge & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 593-606). Oxford: Oxford University Press.
- Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 417-429.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269-293). Oxford: Elsevier.

- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438-443.
- Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, 65(1), 58-71.
- Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, 15(3), 566-573.
- Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Evidence from a scene onset delay paradigm. *Visual Cognition*, 17(6), 1055-1082.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210-228.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113-136.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431-439.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489-1506.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135, 12-35.
- Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, 114(3), 342-356.
- Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, 9(3), 363-386.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28(10), 245-253.

- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578-586.
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 667-682.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205-231.
- Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*, 322-336.
- Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8):20, 1-19.
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382-405.
- Parkhurst, D., Culurciello, E., & Niebur, E. (2000). Evaluating variable resolution displays with visual search: Task performance and eye movements. In A. T. Duchowski (Ed.), *Proceedings of the Eye Tracking Research & Applications Symposium 2000* (pp. 105-109): Association of Computing Machinery.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107-123.
- Pitt, M. A., Myung, I. J., & Zhang, S. B. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*(3), 472-491.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347-356.
- Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21-30.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.
- Rayner, K. (2009a). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.
- Rayner, K. (2009b). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, *2*(5), 1-10.

- Rayner, K., & Duffy, S. A. (1986). Lexical ambiguity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191-201.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47(21), 2714-2726.
- Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, 33A, 351-373.
- Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1188-1200.
- Reddi, B. A. J., Asrress, K. N., & Carpenter, R. H. S. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology*, 90(5), 3538-3546.
- Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, 13(3), 115-119.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125-157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445-526.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2012). Eye movements in reading versus non-reading tasks: Using E-Z Reader to understand the role of word/stimulus familiarity. *Visual Cognition*.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1), 1-21.
- Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive Neuroscience*, 14(3), 371-388.
- Reingold, E. M., & Stampe, D. M. (2003). Using the saccadic inhibition paradigm to investigate saccadic control in reading. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movements* (pp. 347-360). Amsterdam: Elsevier Science.

- Reingold, E. M., & Stampe, D. M. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 194-211.
- Reingold, E. M., Yang, J. M., & Rayner, K. (2010). The time course of word frequency and case alternation effects on fixation times in reading: Evidence for lexical control of eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1677-1683.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21), 9475-9489.
- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2), 113-121.
- Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics*, 25(2), 119-125.
- Schad, D., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*.
- Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, 9(10), 2195-2200.
- Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, 53(3), 305-314.
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4(4), 147-155.
- Sivanandam, S. N., & Deepa, S. N. (2007). *Introduction to genetic algorithms*. Berlin Heidelberg New York: Springer.
- Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1280-1293.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 1-23.
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766-786.

- Trukenbrod, H. A., & Engbert, R. (submitted). ICAT: A computational model for the adaptive control of fixation durations.
- Underwood, G. (2009). Cognitive processes in eye guidance: Algorithms for attention in image processing. *Cognitive Computation*, 1, 64-76.
- van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337-355). Oxford: Elsevier.
- Vergilino-Perez, D., Collins, T., & Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research*, 44(17), 2009-2017.
- Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3):24, 1-15.
- Wang, H.-C., Hwang, A. D., & Pomplun, M. (2010). Object frequency and predictability effects on eye fixation durations in real-world scene viewing. *Journal of Eye Movement Research*, 3(3):3, 1-10.
- Yang, S.-N. (2009). Effects of gaze-contingent text changes on fixation duration in reading. *Vision Research*, 49(23), 2843-2855.
- Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, 41(25-26), 3567-3585.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787-835.
- Zelinsky, G. J. (2012). TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition*.

Acknowledgements

Data collection was supported by a grant from the Economic and Social Research Council (ESRC) of the UK to JMH (RES-062-23-1092). We thank Tim J. Smith for research assistance and discussion.

Footnotes

¹ When modeling the data from the scene onset delay paradigm, we assumed a unidirectional adjustment of fixation duration: difficulties in processing the scene during the delay period inhibit (i.e., delay) saccade initiation, leading to longer fixation durations (Nuthmann, et al., 2010). In contrast, the toy simulations in Figure 2 describe a bidirectional (speed up and slow down) adjustment of fixation durations. Whether one-way or two-way effects occur in scene viewing is an open empirical question.

² Note that these parameters were fit by eye; no advanced fitting procedure was implemented.

Tables

Table 1

Model Parameters for Simulations With the Baseline Model

Model Components	Parameter	Function	Default Value ^a
Saccade timing (random walk)	N	number of states of random-walk timer	11
	t_{sac}	duration of timer interval (ms)	250
Saccade programming	τ_{lab}	labile stage (ms)	180
	τ_{nlab}	non-labile stage (ms)	40
	τ_{ex}	saccade execution (ms)	40

^a Default parameters values are taken from Nuthmann et al. (2010).

Table 2

Mean Fixation Durations (and Standard Errors) in ms For Each Stimulus Onset Delay

Condition Across Tasks

	Stimulus onset delay					
Task	0 ms	300 ms	400 ms	600 ms	800 ms	infinite
Text	184.3	380.1	396.1	443.3	487.3	580.7
reading	(6.8)	(16.7)	(23.9)	(27.3)	(35.0)	(77.5)
Scene	262.1	396.1	440.3	464.9	525.9	462.5
viewing	(5.6)	(22.6)	(34.4)	(35.5)	(50.1)	(53.0)

Table 3

Model Parameters for Modeling of Fixation Durations in the Text Onset Delay and Scene Onset Delay Conditions

Model Components	Parameter	Function	Texts, M	Texts, SD	Scenes, M^a	Scenes, SD^a
Saccade timing (random walk)	$r_1 = \frac{N}{t_{sac}}$	default random walk transition rate	$t_{sac} = 226$ (100–250)	$N = 35$ (5–100)	$t_{sac} = 250$	$N = 17$
	r_0	random walk transition rate during stimulus absence	$r_0 = 0.26 * r_1$		$r_0 = 0.30 * r_1$	
Saccade programming	τ_{lab}	labile stage (ms)	104 (50–150)	$1/3 * M$	180	$1/3 * M$
	τ_{nlab}	non-labile stage (ms)	14 (5–50)	$1/3 * M$	40	$1/3 * M$
	τ_{ex}	saccade execution (ms)	40	$1/3 * M$	40	$1/3 * M$
	$p1_{canc}$	probability of saccade cancellation at stimulus disappearance	0.48		0.5	
	$p2_{canc}$	probability of saccade cancellation at stimulus reappearance	0.42		0.67	

^a Parameters are taken from Nuthmann et al. (2010).

Figure Captions

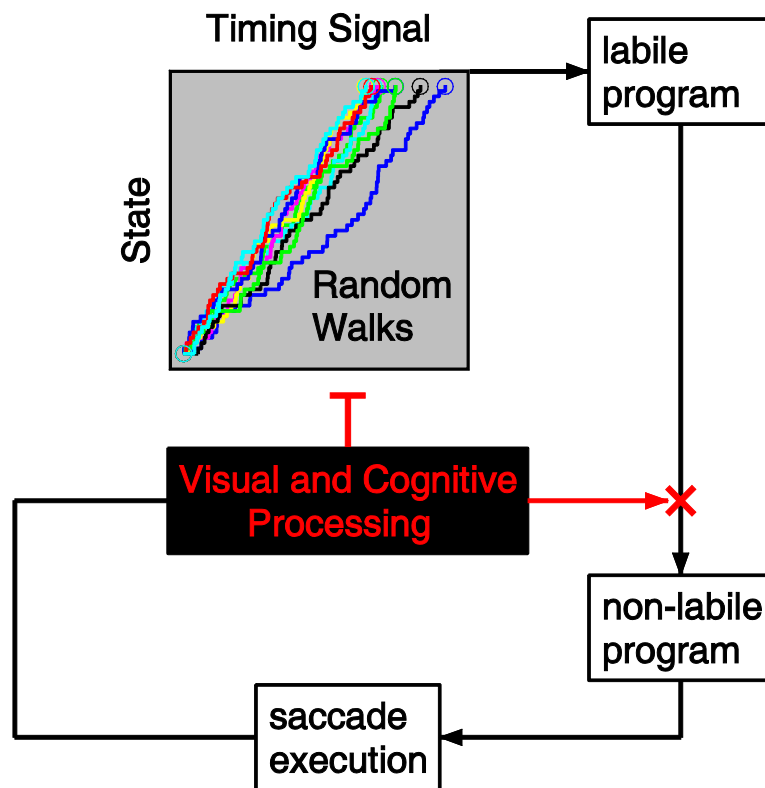


Figure 1. Model overview.

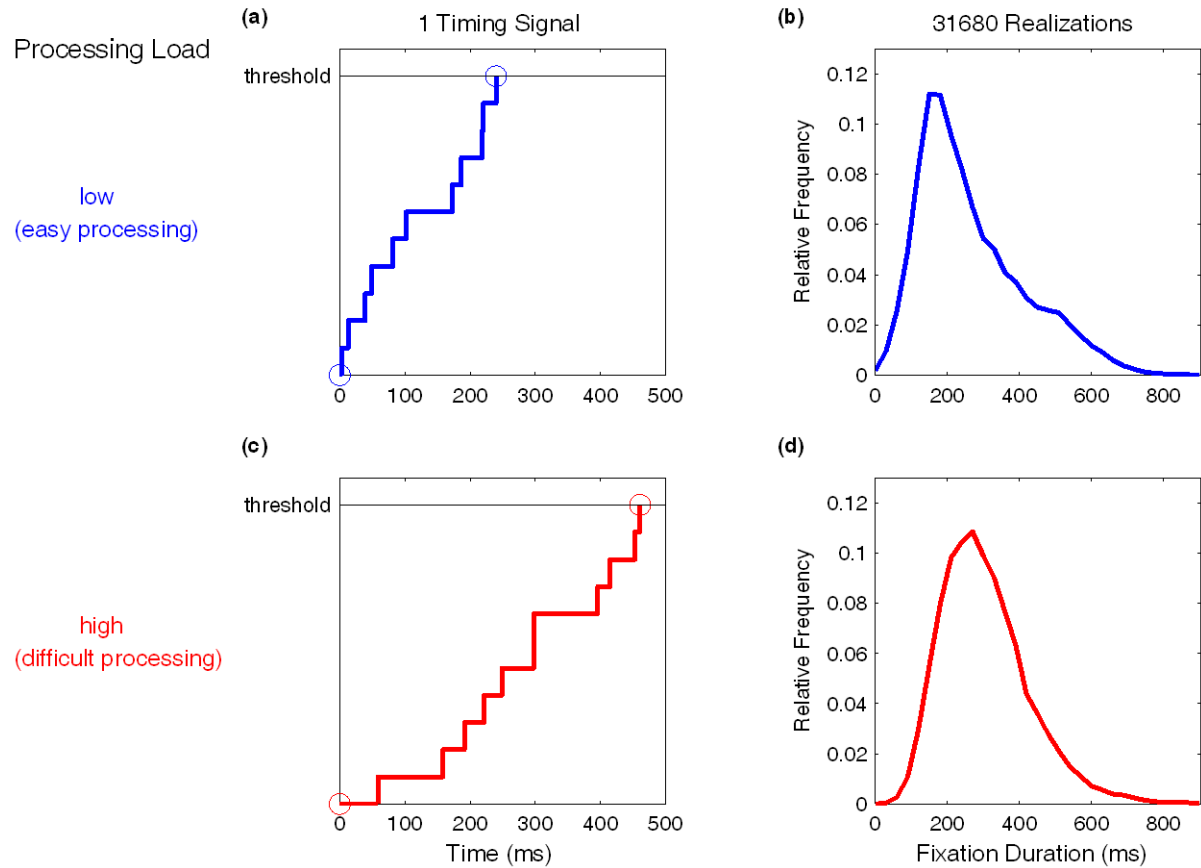


Figure 2. Processing demands modulate the random walk transition rate. Simulations with the baseline model. The two upper plots represent low processing load (“easy processing”) by showing one example timing signal (a) and the obtained fixation duration distribution (b). Accordingly, the two lower plots (c and d) simulate increased processing demands. To adjust for processing difficulty, the default random walk transition rate (cf., Equation 1 and Table 1) was increased (“easy processing”) or decreased (“difficult processing”) by factor 0.80.

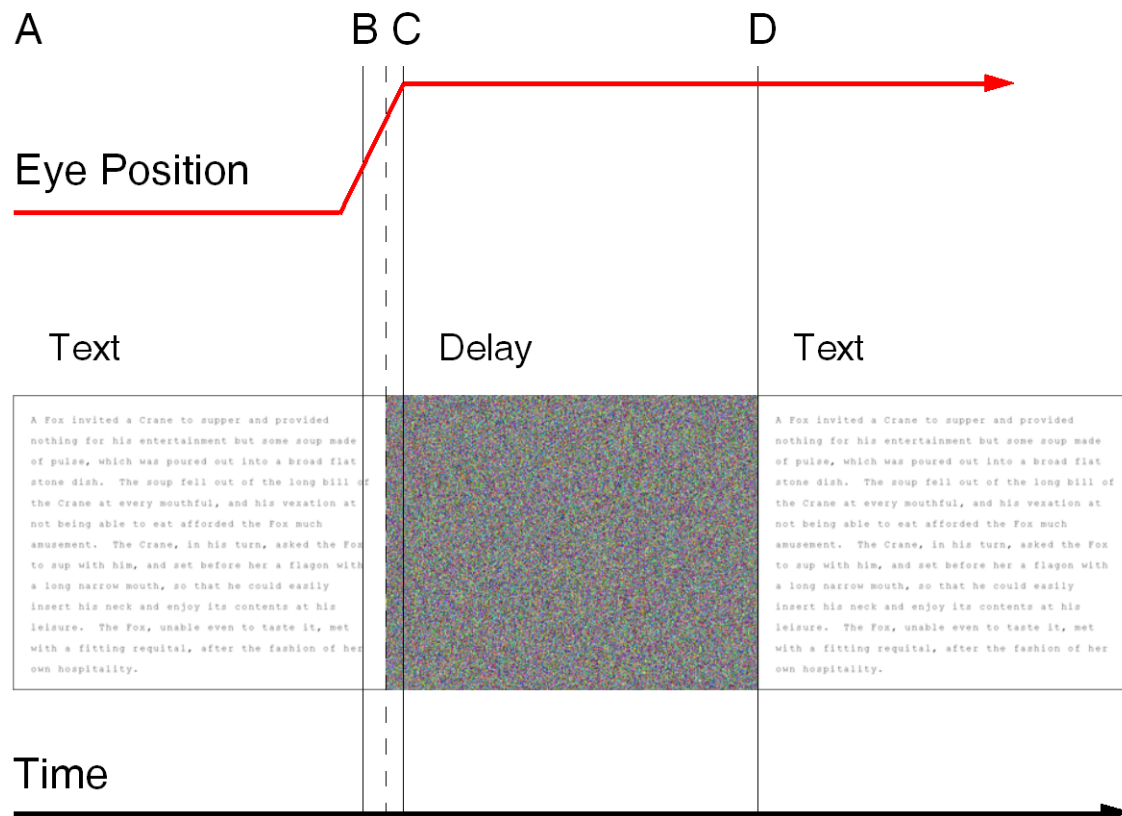


Figure 3. Illustration of the stimulus onset paradigm for the text reading condition. At the first time-point (A), the eyes are in the fifth fixation following the last delay. After the next saccade is detected (B), the display is changed (vertical broken line), so that when the eyes begin the following critical fixation (C), the text has been removed from view. Following the specified delay (D), the text reappears. The duration of the remaining saccade following the display change (C - B) is subtracted from the specified delay (D - B) to generate the actual delay. Texts and mask were presented in color.

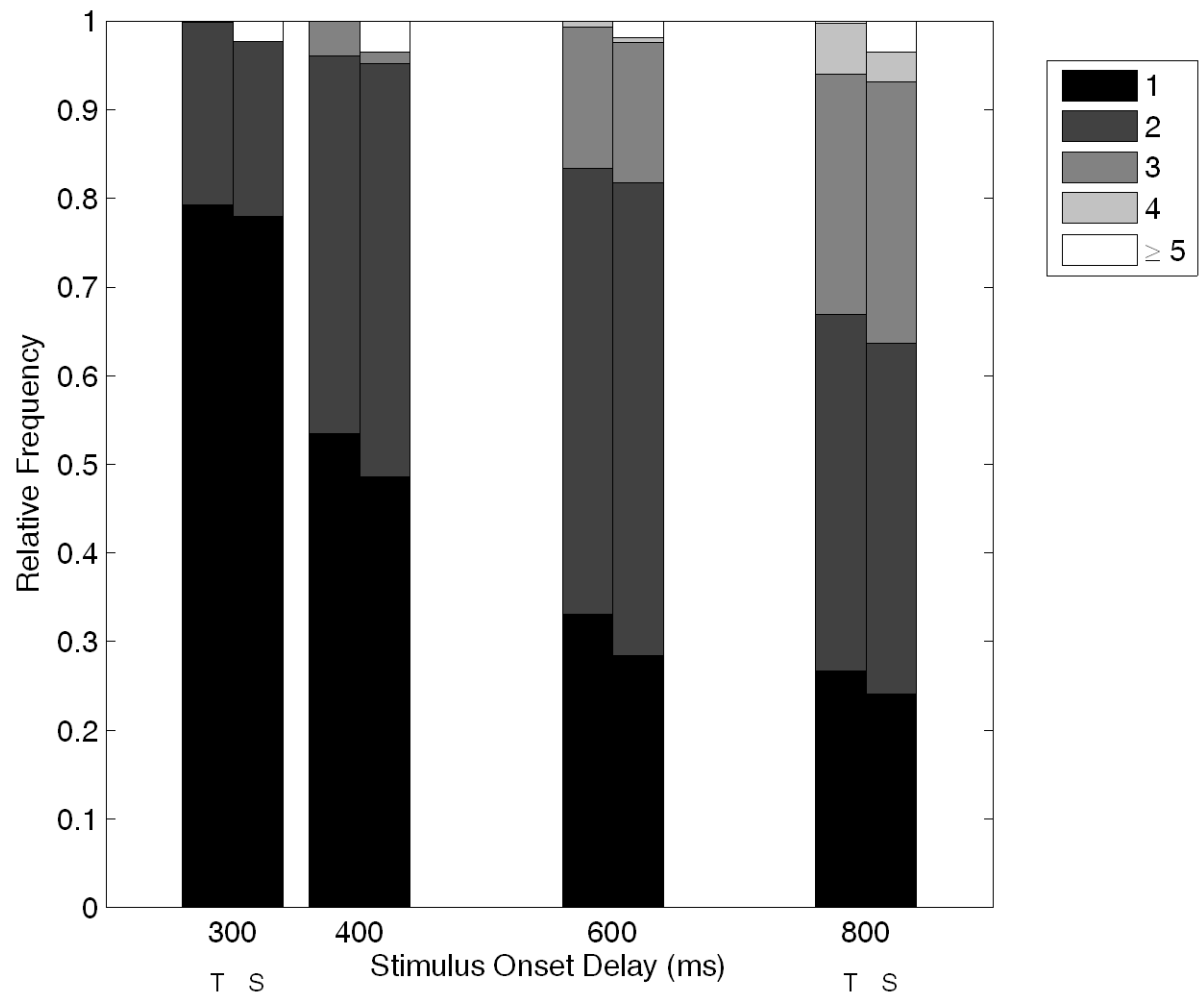


Figure 4. Total number of fixations during stimulus absence as a function of stimulus onset delay and task. For each stimulus onset delay incident it was determined how many fixations were made during stimulus absence. The stacked bars represent the relative frequencies of 1-, 2-, 3-, 4-, and ≥ 5 -fixation cases. For each delay condition, plotted on the x -axis, the left bar (T) represents the text reading data and the right bar (S) the scene viewing data.

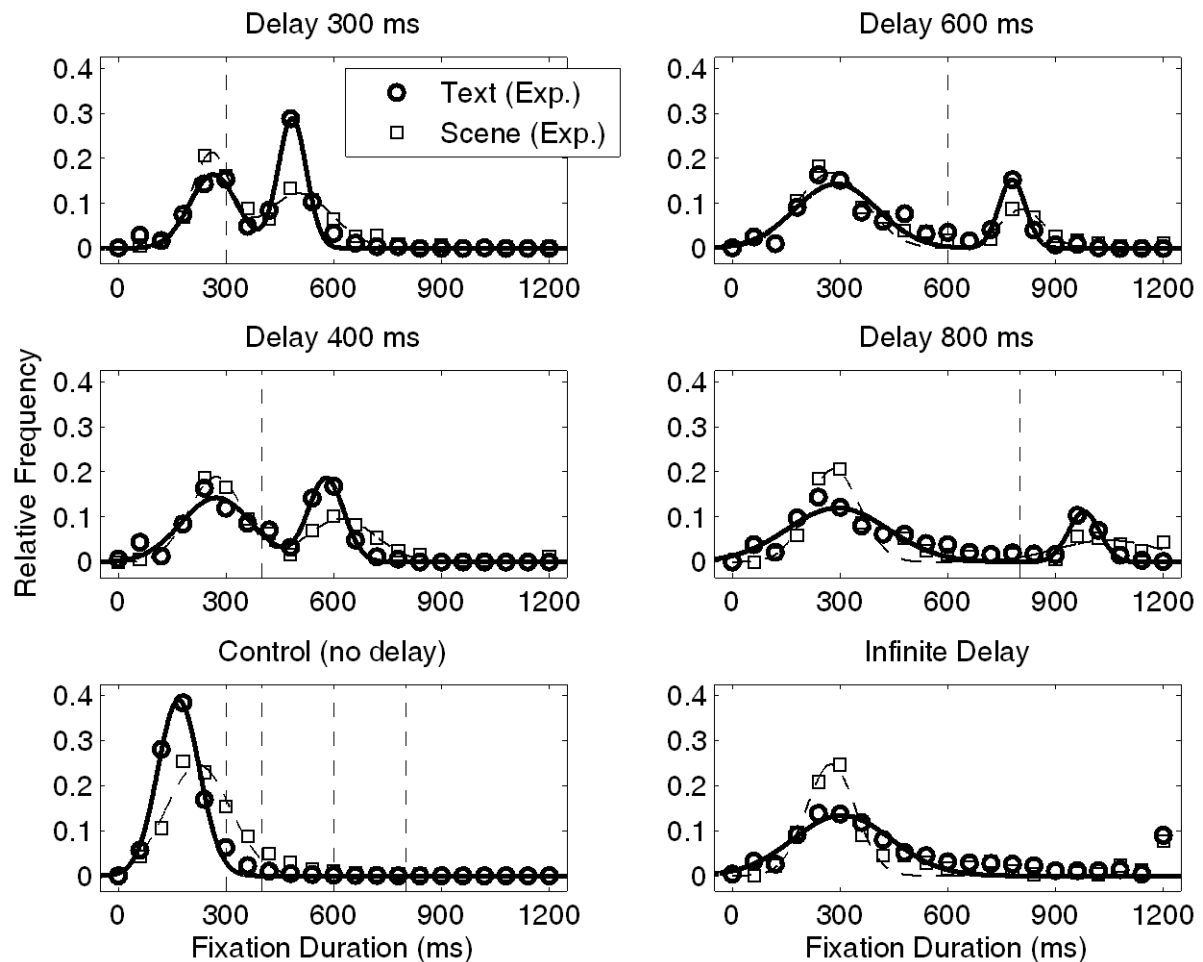


Figure 5. Distribution functions of fixation duration in the stimulus onset delay experiment. Frequency of occurrence is calculated for each 60-ms bin. In a given panel, the empirical text reading data (bold circles) are compared with scene viewing data (squares). In addition, one-term or two-term Gaussian models were fit to the data (text data: bold solid line, scene data: broken line). Vertical broken lines mark the delay duration. Note that in all (non-infinite) delay conditions, the second peak of the distribution rises about 100 ms following stimulus reinstatement.

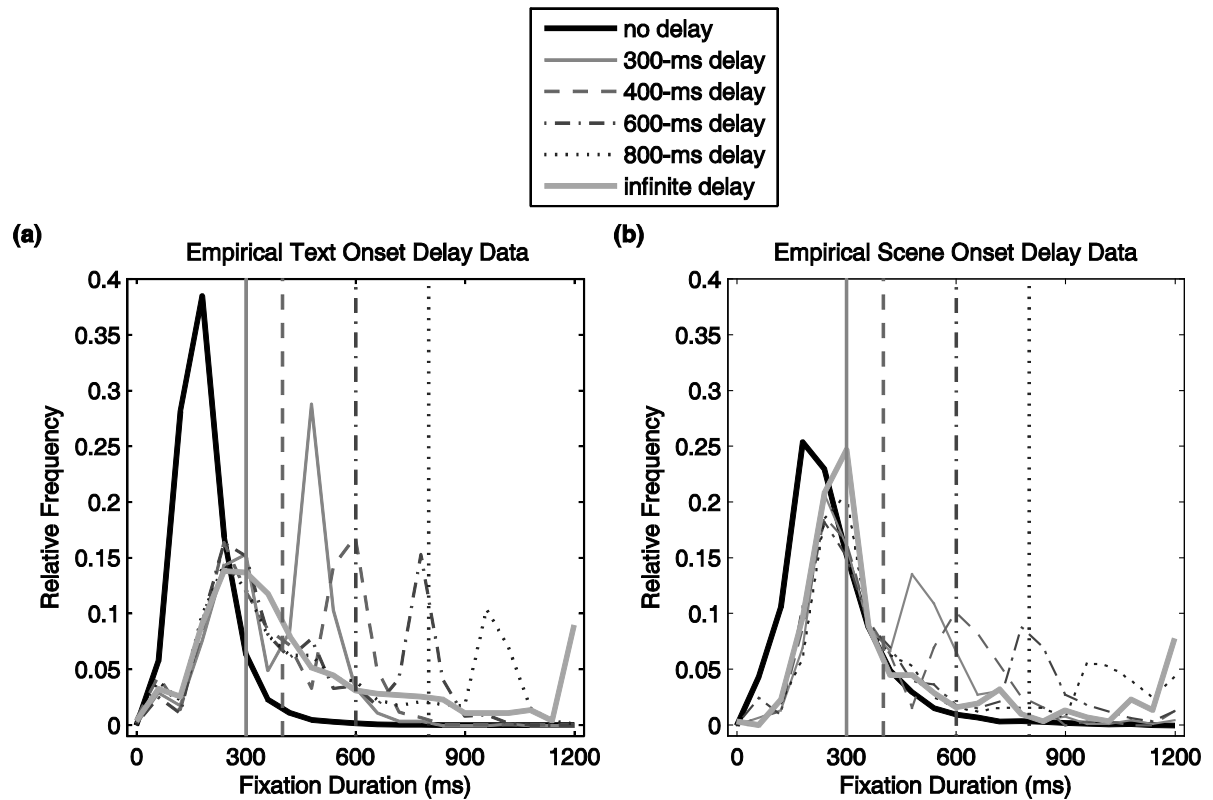


Figure 6. Distribution functions of fixation duration in the stimulus onset delay experiment. Data from Figure 5 are replotted to allow direct comparison of distributions for the different onset delay conditions and the 0-ms delay control condition for text reading (a) and scene viewing (b). For all delay conditions but the infinite delay, vertical lines mark the delay duration.

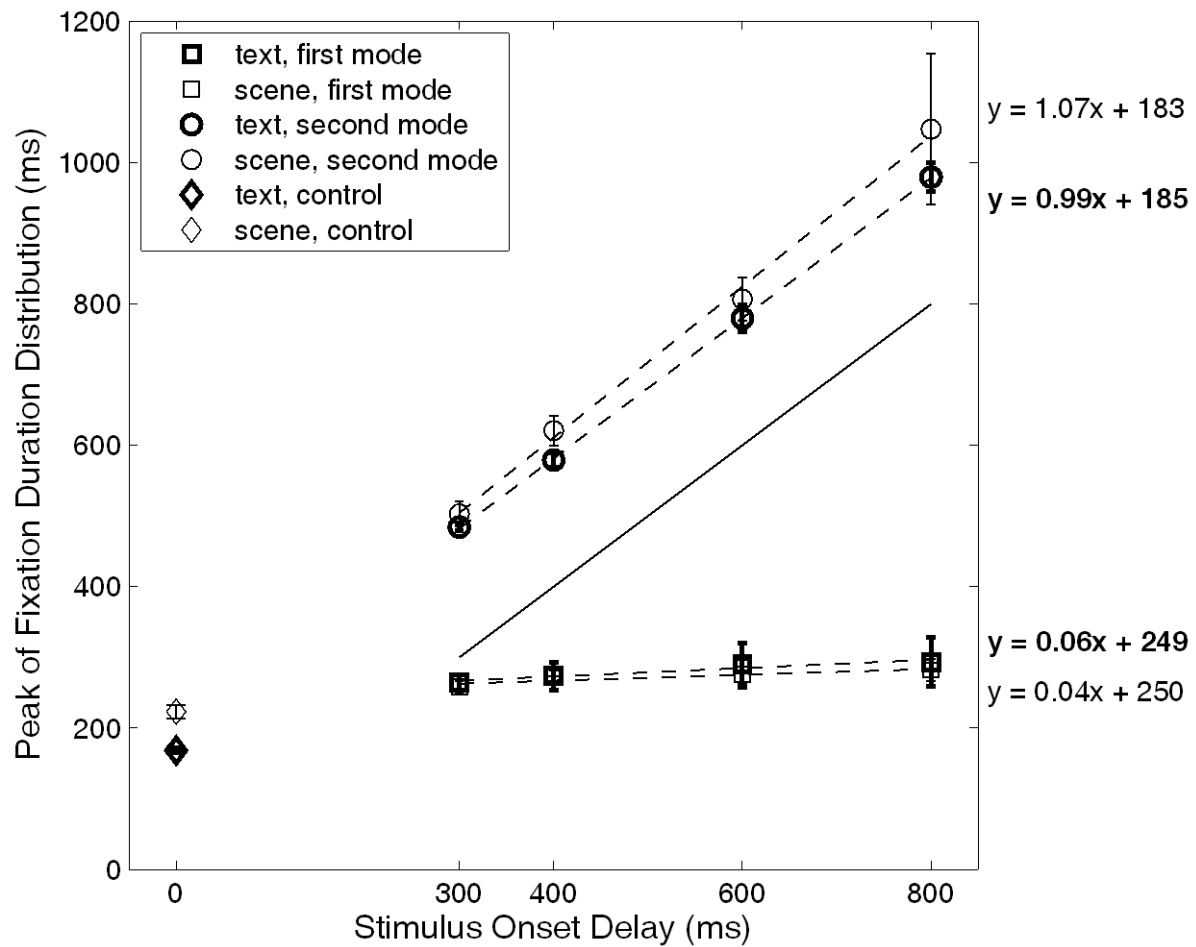


Figure 7. Modes of critical fixation duration distributions. The distribution mode in the 0-ms delay control condition is contrasted with the bimodal critical fixation duration distributions in the different delay conditions. One-term or two-term Gaussian models were fit to the data (see Figure 5). For each stimulus onset delay, the point below the diagonal displays the first, early mode of the distribution, and the point above the diagonal the second, late mode. Error bars represent 95% confidence intervals for the coefficient estimates. Linear regressions were fit to the first and second modes across the stimulus onset delay conditions; the figure shows the best fitting regression lines and corresponding equations. Bold symbols and equations represent text reading data.

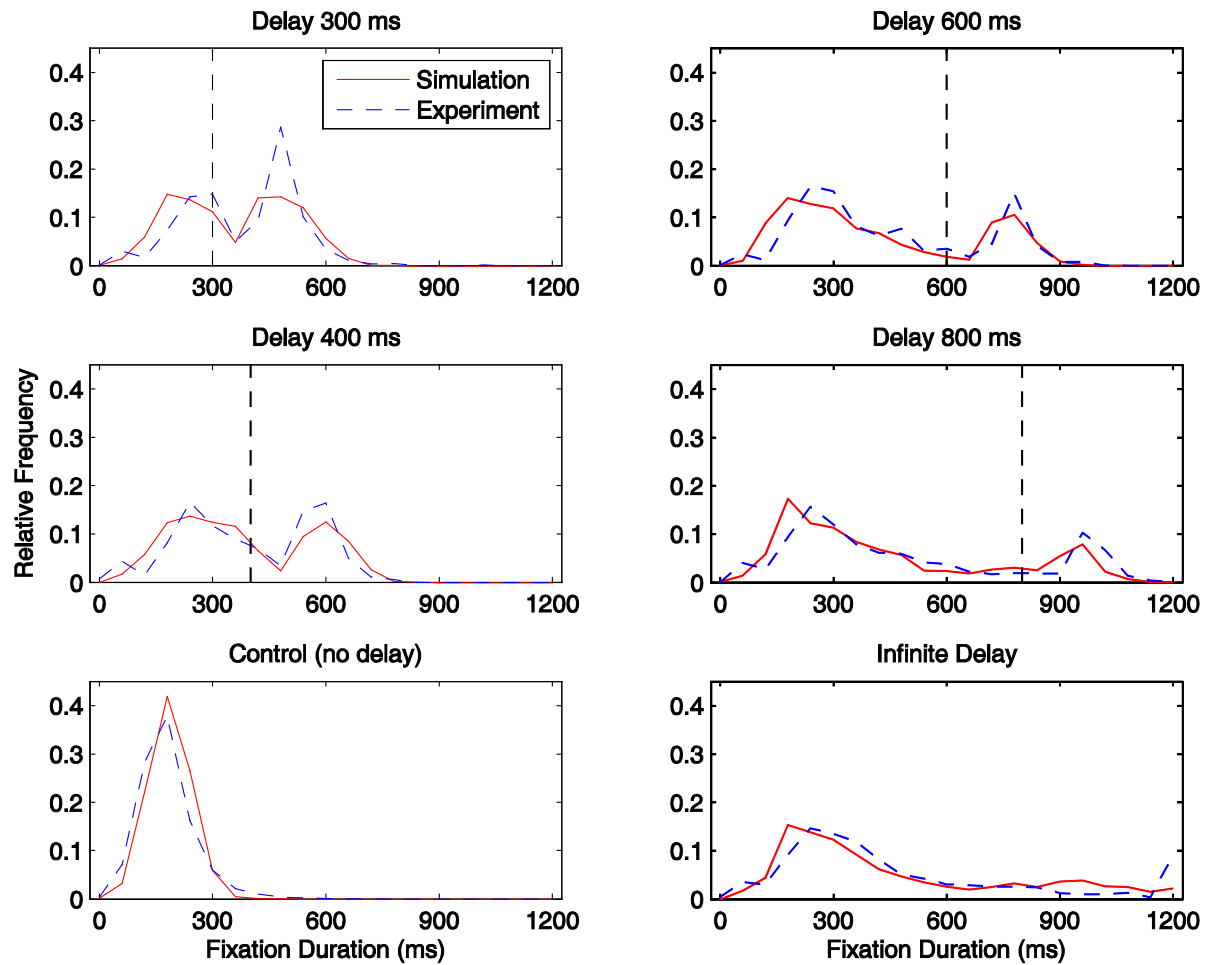


Figure 8. Distribution functions of fixation duration in the text onset delay condition.

Frequency of occurrence is calculated for each 60-ms bin. In a given panel, simulated data (solid line) are compared with empirical data (broken line). Vertical broken lines mark the delay duration.